Math 164: Problem Set 7

Owen Jones

2/23/2024

9.1 (a) $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \mathbf{F}(\mathbf{x}^{(k)})^{-1}\mathbf{g}^{(k)}$
$= \mathbf{x}^{(k)} - \frac{4(\mathbf{x}^{(k)}-\mathbf{x_0})^3}{12(\mathbf{x}^{(k)}-\mathbf{x_0})^2}$
$\Rightarrow \mathbf{x}^{(k+1)} = \frac{2}{3}\mathbf{x}^{(k)} + \frac{1}{3}\mathbf{x_0}$
subtracting $\mathbf{x_0}$ from both sides $\mathbf{x}^{(k+1)} - \mathbf{x_0} = \frac{2}{3}(\mathbf{x}^{(k)} - \mathbf{x_0})$

(b) From part (a), $\mathbf{y}^{(k+1)} = |\mathbf{x}^{(k+1)} - \mathbf{x_0}| = \frac{2}{3}|\mathbf{x}^{(k)} - \mathbf{x_0}| = \frac{2}{3}\mathbf{y}^{(k)}$, so the sequence satisfies $\mathbf{y}^{(k+1)} = \frac{2}{3}\mathbf{y}^{(k)}$.

(c) From part (b) we have $\mathbf{y}^{(k+1)} = \frac{2}{3}\mathbf{y}^{(k)} \Rightarrow \mathbf{y}^{(k)} \to 0$. Hence, $\mathbf{x}^{(k)} \to \mathbf{x_0}$ for any $\mathbf{x}^{(0)}$.

(d) From part (c), $\mathbf{x}^{(k)} \to \mathbf{x_0}$ for any $\mathbf{x}^{(0)}$. From part (b), $\lim_{k\to\infty} \dfrac{|\mathbf{x}^{(k+1)} - \mathbf{x_0}|}{|\mathbf{x}^{(k)} - \mathbf{x_0}|} = \lim_{k\to\infty} \dfrac{\mathbf{y}^{(k+1)}}{\mathbf{y}^{(k)}} = \dfrac{2}{3} > 0$. Thus, the order of convergence is linear.

(e) The theorem assumes $\mathbf{F}(\mathbf{x}^*)$ is invertible, but $\mathbf{x}^* = \mathbf{x_0}$.
Thus, $F(\mathbf{x}^*) = 12(\mathbf{x}^* - \mathbf{x_0})^2 = 0$, so $\mathbf{F}(\mathbf{x}^*)$ is singular.

9.3 (a) $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \mathbf{F}(\mathbf{x}^{(k)})^{-1}\mathbf{g}^{(k)}$
$= \mathbf{x}^{(k)} - \frac{\frac{4}{3}(\mathbf{x}^{(k)})^{\frac{1}{3}}}{\frac{4}{9}(\mathbf{x}^{(k)})^{\frac{-2}{3}}} = -2\mathbf{x}^{(k)}$

(b) From part (a) we have $\mathbf{x}^{(k)} = (-2)^k\mathbf{x_0}$ which diverges for any $\mathbf{x_0} \neq 0$.

9.4 (a) $f(x) \geq 0 \ \forall x \in \mathbb{R}^2$, so it suffices to show $x = [1,1]^\top \Leftrightarrow f(x) = 0$.
$f([1,1]^\top) = 0$ by plugging in numbers
$f(x) = 0 \Leftrightarrow x_2 - x_1^2 = 0, (1-x_1)^2 = 0 \Rightarrow x_1 = 1, x_2 = x_1^2 \Rightarrow x = [1,1]^\top$

(b)

$$\nabla f(x) = [400x_1^3 - 400x_1x_2 + 2(x_1-1), 200(x_2 - x_1^2)]^\top$$
$$F(x) = \begin{bmatrix} 1200x_1^2 - 400x_2 + 2 & -400x_1 \\ -400x_1 & 200 \end{bmatrix}$$
$$\Rightarrow F(x)^{-1} = \frac{1}{80000(x_1^2 - x_2) + 400} \begin{bmatrix} 200 & 400x_1 \\ 400x_1 & 1200x_1^2 - 400x_2 + 2 \end{bmatrix}$$
$$\mathbf{x}^{(1)} = \mathbf{x}^{(0)} - \frac{1}{400}\begin{bmatrix} 200 & 0 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} -2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$
$$\mathbf{x}^{(2)} = \mathbf{x}^{(1)} - \frac{1}{80400}\begin{bmatrix} 200 & 400 \\ 400 & 1202 \end{bmatrix}\begin{bmatrix} 400 \\ -200 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

(c)

$$\mathbf{x}^{(1)} = \mathbf{x}^{(0)} - 0.05 \cdot \begin{bmatrix} -2 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.1 \\ 0 \end{bmatrix}$$
$$\mathbf{x}^{(2)} = \mathbf{x}^{(1)} - 0.05 \cdot \begin{bmatrix} -1.4 \\ -2 \end{bmatrix} = \begin{bmatrix} 0.17 \\ 0.1 \end{bmatrix}$$

9.5 Because the case $\mathbf{x}_0 = \mathbf{x}^*$ is trivial, assume $\mathbf{x}_0 \neq \mathbf{x}^*$
From standard Newton's method min $f(\mathbf{x}) = f(\mathbf{x}^*) = f(\mathbf{x}_0 - \mathbf{F}(\mathbf{x}_0)^{-1}\mathbf{g}^{(k)})$.
It follows $f(\mathbf{x}_0 - \mathbf{F}(\mathbf{x}_0)^{-1}\mathbf{g}^{(k)}) \leq f(\mathbf{x}_0 - \alpha\mathbf{F}(\mathbf{x}_0)^{-1}\mathbf{g}^{(k)})\ \forall\alpha \geq 0$
Thus, $alpha_0 = \underset{\alpha\geq 0}{\arg\min} f(\mathbf{x}_0 - \alpha\mathbf{F}(\mathbf{x}_0)^{-1}\mathbf{g}^{(k)}) = 1$.
Hence, $f(\mathbf{x}_0 - \alpha_0\mathbf{F}(\mathbf{x}_0)^{-1}\mathbf{g}^{(k)})$ is equivalent to the standard Newton algorithm, so it also converges in a single step.

10.1 We will show by induction that the set $\{\mathbf{d}^{(0)}, \ldots, \mathbf{d}^{(n-1)}\}$ is $\mathbf{Q}$-conjugate.
The base case $k = 0$ is trivial because there is only one vector in the set.
Induction Hypothesis: Assume for some $k < n-1$ the set $\{\mathbf{d}^{(0)}, \ldots, \mathbf{d}^{(k)}\}$ is $\mathbf{Q}$-conjugate.
Induction step: Fix some $j = 0 \ldots k$. WTS $\mathbf{d}^{(k+1)^\top}\mathbf{Q}\mathbf{d}^{(j)} = 0$
$$\mathbf{d}^{(k+1)^\top}\mathbf{Q}\mathbf{d}^{(j)} = \mathbf{p}^{(k+1)^\top}\mathbf{Q}\mathbf{d}^{(j)} - \sum_{i=1}^{k}\frac{\mathbf{p}^{(k+1)^\top}\mathbf{Q}\mathbf{d}^{(i)}}{\mathbf{d}^{(i)^\top}\mathbf{Q}\mathbf{d}^{(i)}}\mathbf{d}^{(i)}\mathbf{Q}\mathbf{d}^{(j)}$$
By induction hypothesis: $\{\mathbf{d}^{(0)}, \ldots, \mathbf{d}^{(k)}\}$ is $\mathbf{Q}$-conjugate, so $\mathbf{d}^{(i)}\mathbf{Q}\mathbf{d}^{(j)}$ $\forall i \neq j$.
Thus $\mathbf{d}^{(k+1)^\top}\mathbf{Q}\mathbf{d}^{(j)} = \mathbf{p}^{(k+1)^\top}\mathbf{Q}\mathbf{d}^{(j)} - \frac{\mathbf{p}^{(k+1)^\top}\mathbf{Q}\mathbf{d}^{(j)}}{\mathbf{d}^{(j)^\top}\mathbf{Q}\mathbf{d}^{(j)}}\mathbf{d}^{(j)}\mathbf{Q}\mathbf{d}^{(j)}$
$= \mathbf{p}^{(k+1)^\top}\mathbf{Q}\mathbf{d}^{(j)} - \mathbf{p}^{(k+1)^\top}\mathbf{Q}\mathbf{d}^{(j)} = 0$
Hence, by induction, the set $\{\mathbf{d}^{(0)}, \ldots, \mathbf{d}^{(k)}\}$ is $\mathbf{Q}$-conjugate.

10.4 (a) Because $\mathbf{Q}$ is symmetric, there exists an orthogonal eigenbasis $\{\mathbf{d}^{(1)}, \ldots, \mathbf{d}^{(n)}\}$ for $\mathbf{Q}$ with real eigenvalues. It follows for any $i \neq j$ $\mathbf{d}^{(i)^\top}\mathbf{Q}\mathbf{d}^{(j)} = \lambda_j\mathbf{d}^{(i)^\top}\mathbf{d}^{(j)} = 0$. Thus, $\{\mathbf{d}^{(1)}, \ldots, \mathbf{d}^{(n)}\}$ is $\mathbf{Q}$-conjugate.

(b) Let $\lambda_i := \frac{\mathbf{d}^{(i)^\top}\mathbf{Q}\mathbf{d}^{(i)}}{\mathbf{d}^{(i)^\top}\mathbf{d}^{(i)}}$. $\{\mathbf{d}^{(1)}, \ldots, \mathbf{d}^{(n)}\}$ must be linearly independent because the set is $\mathbf{Q}$-conjugate. Let $\mathbf{D} = \begin{bmatrix} \mathbf{d}^{(1)^\top} \\ \vdots \\ \mathbf{d}^{(n)^\top} \end{bmatrix}$ which is invertible because $\{\mathbf{d}^{(1)}, \ldots, \mathbf{d}^{(n)}\}$ is linearly independent. Because the set $\{\mathbf{d}^{(1)}, \ldots, \mathbf{d}^{(n)}\}$ is assumed to be orthogonal, $\lambda_i \mathbf{d}^{(i)^\top}\mathbf{d}^{(j)} = 0$ $\forall i \neq j$. Because the set $\{\mathbf{d}^{(1)}, \ldots, \mathbf{d}^{(n)}\}$ is $\mathbf{Q}$-conjugate, we have $\mathbf{d}^{(i)^\top}\mathbf{Q}\mathbf{d}^{(j)} = 0$ $\forall i \neq j$. Thus $\lambda_i \mathbf{d}^{(i)^\top}\mathbf{d}^{(j)} = \mathbf{d}^{(i)^\top}\mathbf{Q}\mathbf{d}^{(j)}$. By how we defined λ_i, we have $\lambda_i \mathbf{d}^{(i)^\top}\mathbf{d}^{(i)} = \mathbf{d}^{(i)^\top}\mathbf{Q}\mathbf{d}^{(i)}$. It follows $\mathbf{D}\mathbf{Q}\mathbf{d}^{(i)} = \mathbf{D}(\lambda_i \mathbf{d}^{(i)})$ $\forall i = 1 \ldots n$ because

$$\begin{bmatrix} \mathbf{d}^{(1)^\top}\mathbf{Q}\mathbf{d}^{(i)} \\ \vdots \\ \mathbf{d}^{(n)^\top}\mathbf{Q}\mathbf{d}^{(i)} \end{bmatrix} = \begin{bmatrix} \lambda_i \mathbf{d}^{(1)^\top}\mathbf{d}^{(i)} \\ \vdots \\ \lambda_i \mathbf{d}^{(n)^\top}\mathbf{d}^{(i)} \end{bmatrix}.$$

Because $\mathbf{D}$ is invertible, we have $\mathbf{Q}\mathbf{d}^{(i)} = \lambda_i \mathbf{d}^{(i)}$ $\forall i = 1 \ldots n$. Hence each vector in the set $\{\mathbf{d}^{(1)}, \ldots, \mathbf{d}^{(n)}\}$ is an eigenvector of $\mathbf{Q}$.

10.5 Premultiplying both sides by $\mathbf{d}^{(k)^\top}\mathbf{Q}$ we obtain
$\mathbf{d}^{(k)^\top}\mathbf{Q}\mathbf{d}^{(k+1)} = \gamma_k \mathbf{d}^{(k)^\top}\mathbf{Q}\mathbf{g}^{(k+1)} + \mathbf{d}^{(k)^\top}\mathbf{Q}\mathbf{d}^{(k)}$.
Using the fact $\mathbf{d}^{(k)}$ and $\mathbf{d}^{(k+1)}$ are $\mathbf{Q}$-conjugate $\mathbf{d}^{(k)^\top}\mathbf{Q}\mathbf{d}^{(k+1)} = 0$.
Thus, $\gamma_k = -\frac{\mathbf{d}^{(k)^\top}\mathbf{Q}\mathbf{d}^{(k)}}{\mathbf{d}^{(k)^\top}\mathbf{Q}\mathbf{g}^{(k+1)}}$

10.7 $\phi(\alpha) = \frac{1}{2}(\mathbf{x}_0 + \mathbf{D}\alpha)^\top \mathbf{Q}(\mathbf{x}_0 + \mathbf{D}\alpha) - (\mathbf{x}_0 + \mathbf{D}\alpha)^\top \mathbf{b}$
$= \frac{1}{2}\mathbf{x}_0^\top \mathbf{Q}\mathbf{x}_0 + \alpha^\top \mathbf{D}^\top \mathbf{Q}\mathbf{x}_0 + \frac{1}{2}\alpha^\top \mathbf{D}^\top \mathbf{Q}\mathbf{D}\alpha - \alpha^\top \mathbf{D}^\top \mathbf{b} - \mathbf{x}_0^\top \mathbf{b}$
$= \frac{1}{2}\alpha^\top \mathbf{D}^\top \mathbf{Q}\mathbf{D}\alpha - \alpha^\top(\mathbf{D}^\top \mathbf{b} - \mathbf{D}^\top \mathbf{Q}\mathbf{x}_0) + (\frac{1}{2}\mathbf{x}_0^\top \mathbf{Q}\mathbf{x}_0 - \mathbf{x}_0^\top \mathbf{b})$.
Since we have shown $\phi(\alpha)$ is a quadratic function on $\mathbb{R}^r$, it suffices to show $\mathbf{D}^\top \mathbf{Q}\mathbf{D} > 0$.
Since $\mathbf{Q} > 0$, $\alpha^\top \mathbf{D}^\top \mathbf{Q}\mathbf{D}\alpha = (\mathbf{D}\alpha)^\top \mathbf{Q}(\mathbf{D}\alpha) \geq 0$, and $\alpha^\top \mathbf{D}^\top \mathbf{Q}\mathbf{D}\alpha = 0$ iff $\mathbf{D}\alpha = 0$.
Since $rank(\mathbf{D}) = r$ and $\alpha \in \mathbb{R}^r$, $\mathbf{D}\alpha = 0$ iff $\alpha = 0$. Thus, $\alpha^\top \mathbf{D}^\top \mathbf{Q}\mathbf{D}\alpha = 0$ iff $\alpha = 0$. Thus, $\mathbf{D}^\top \mathbf{Q}\mathbf{D} > 0$.